

WORLD METEOROLOGICAL ORGANIZATION

INTERGOVERNMENTAL OCEANOGRAPHIC  
COMMISSION (OF UNESCO)

---

JOINT WMO/IOC TECHNICAL COMMISSION FOR  
OCEANOGRAPHY AND MARINE METEOROLOGY (JCOMM)  
EXPERT TEAM ON MARINE CLIMATOLOGY

---

ETMC-II/Doc. 4.1  
(15.II.2007)

SECOND SESSION

---

ITEM 4.1

GENEVA, SWITZERLAND, 26 TO 27 MARCH 2007

Original: ENGLISH

## DATA ARCHIVAL

### Development of International Marine Meteorological Archive (IMMA) format

*(Submitted by Mr Scott Woodruff<sup>1</sup>, Ms Sandy Lubker<sup>1</sup>, Mr Steve Worley<sup>2</sup> and Mr Zaihua Ji<sup>2</sup>)*

*(1) NOAA Earth System Research Laboratory (ESRL), USA*

*(2) National Center for Atmospheric Research (NCAR), USA*

---

### Summary and purpose of document

This document reviews the background of the International Maritime Marine Meteorological Archive (IMMA) format, and updates its status - including widespread public availability for climate research of historical and contemporary marine data and metadata in (from the periods of 1784-2005), and Auxiliary (1750 to present) to, the International Comprehensive Ocean-Atmosphere Data Set (ICOADS).

---

### ACTION PROPOSED

The Expert Team on Marine Climatology is invited to:

- (a) Consider and discuss the background and status of the IMMA format, including possible broadened technical review within the ETMC and by its proposed new cross-cutting Task Team on Delayed-Mode VOS data (TT-DMVOS; ref. ETMC-II/Doc. 3.1);
- (b) Agree on any additional future plans for the ETMC regarding this issue, and also on any possible recommendations to the JCOMM and other appropriate WMO bodies aimed to foster utilization and future development of the format - these actions could potentially extend to a JCOMM publication, or consideration of the feasibility of more formalized adoption.

---

**Appendix:** A. Archival of Data Other Than in IMMT Format: The International Maritime Meteorological Archive (IMMA) Format

## DISCUSSION

### 1. Background

The International Maritime Meteorological Archive (IMMA) format has been under consideration by the Expert Team on Marine Climatology (ETMC), and its predecessor Sub-group under the Commission for Marine Meteorology (CMM), for several years (JCOMM 2000, 2004). The Data Management plan from the First Session of JCOMM (JCOMM-I, Akureyri, Iceland, 19-29 June 2001), tasked the ETMC to finalize the format, with a view to eventual submission to the Commission for formal adoption, and the possibility of its publication through a JCOMM Technical Report has been suggested, most recently by the Second Session of the JCOMM Data Management Coordination Group (DMCG-II, Geneva, Switzerland, 10-12 October 2006).

Further, the Second JCOMM Workshop on Advances in Marine Climatology (CLIMAR-II, Brussels Belgium, 2003) (Parker, et al., 2004), and the First and Second Workshops on Advances in the Use of Historical Marine Climate Data (MARCDAT, Boulder USA, 2002, and Exeter UK, 2005) (Diaz, et al., 2002 and Kent, et al., 2007b), have all suggested continued usage and expansion of the format (recent recommendations from these workshops relating to IMMA are discussed in Section 5).

Following its introduction at JCOMM (2000), the IMMA format and documentation underwent a number of updates and revisions, and a provisional version of the format became fully-operational for the International Comprehensive Ocean-Atmosphere Data Set (ICOADS; Worley, et al., 2005). This discussion, which updates the corresponding document presented at the ETMC-I (JCOMM 2004), provides a brief overview of the format, of its status and current utilization for the ICOADS, for other projects such as the Climatological Database for the World's Oceans 1750-1850 (CLIWOC) (García-Herrera, et al., 2005) and the Voluntary Observing Ship (VOS) Climate (VOSCLIM) Project. Appendix A to this document provides a detailed description and additional background information regarding the format.

### 2. Format overview

The IMMA is a flexible and extensible ASCII format designed for archival and exchange - of both historical and contemporary marine data - plus convenient data access by users. The convention adopted a standardized representation of data within the format uses a "fixed-field" approach (in contrast to a delimited format, or a binary format such as BUFR) similar to IMMT. In addition, blanks fully occupying any field width, define the universal representation for missing data.

The fields presently composing IMMA have been organized into two different types of format components: (i.) the "core", which contains the most universal and commonly used marine data elements (e.g., reported time and location, temperatures, wind, pressure, cloudiness, and sea state) and (ii.) "attachments" (attn), which contain additional, less universal data elements. The "core" forms the common front-end for all IMMA data, and by itself forms a relatively concise "record type" that can satisfy many end-user requirements. Additional record types are constructed by appending to the core one or more attms (Table 1).

Table 1. Size (bytes) and status of IMMA format components: the core and presently defined or proposed attachments (attn).

<u>Abbreviation</u>	<u>Name</u>	<u>Size (B)</u>	<u>Cumulative size (B)</u>	<u>Status</u>
C0	Core	108	108	Operational
C1	ICOADS attn	65	173	"
C2	IMMT/FM 13 attn	76	249	"
C3	Model quality control (QC) attn	66	315	"
C4	Ship metadata attn	57	372	"
C5	Historical attn	(TBD)	(n/a)	under development
C6	Supplemental data attn	(variable)	(n/a)	Operational

Following is a further description of the attachments listed in Table 1:

- C1 (ICOADS): used to store data and metadata fields (e.g., QC flags) specific to ICOADS;
- C2 (IMMT-2/FM 13): stores fields, in addition to those included in the core, from contemporary (e.g., IMMT and FM 13 GTS) ship or other (e.g., drifting and moored buoy) marine data. Note: this attm has not been updated for the IMMT-3;
- C3 (Model QC): designed specifically for the VOSCLim to store co-located numerical weather prediction feedbacks from the United Kingdom Met Office;
- C4 (Ship metadata): stores selected platform and instrumental metadata from WMO (1955–) Publication No. 47 (Kent, et al., 2007a);
- C5 (Historical; still under development): intended for storage of fields from older ship data that are no longer part of the current ship code (e.g., Beaufort force numbers; Woodruff; et al.; 2005);
- C6 (Supplementary data): used to store supplementary (or original input) data of indeterminate type, and of fixed or variable length (e.g., FM 13 strings).

The format has been designed with the flexibility to define new attms, as required, to satisfy emerging or unanticipated data and metadata requirements. For example, a new attm designed to store additional buoy data elements from the GTS FM 18 BUOY format, or buoy metadata, might prove useful in the future.

### 3. Status and utilization

The provisional version of the IMMA format is already widely used for the ICOADS, and also in use for VOSCLim data at that project's Data Assembly Center (DAC), based at the NOAA National Climatic Data Center. In addition, a variety of contemporary and historical data collections have been converted into the IMMA format, generally including the standard ICOADS QC flags, which are offered as datasets "Auxiliary" to the ICOADS (Table 2).

Storing the QC'd Auxiliary data in IMMA format alongside the regular ICOADS data, makes all these data readily accessible to users via web-based subsetting software (available through the project web portal: [icoads.noaa.gov](http://icoads.noaa.gov)). Temporal, spatial, and IMMA content (i.e., individual fields) can be selected through this processing for output into simple ASCII and spreadsheet-compatible formats.

Table 2. ICOADS or Auxiliary datasets currently available in IMMA format. The Auxiliary datasets are divided into two types: "new" sources generally are unique observations not found in ICOADS, whereas "replacement" sources are represented in the ICOADS either by an earlier version or another source that is very similar. These data cannot be added to the ICOADS without careful consideration of the impact of duplicate and near-duplicate records. Total numbers of reports in  $10^6$  (M) or  $10^3$  (K) are listed, and sizes (uncompressed) in megabytes or gigabytes.

	<i>Type</i>	<i>Date Range</i>	<i>Description</i>	<i>No. reports</i>	<i>Size</i>
ICOADS	—	1784-2005	ICOADS Release 2.3 <sup>1</sup>	213 M	49.0 GB
Auxiliary	New	1750-1855	CLIWOC <sup>2</sup>	280 K	148 MB
Auxiliary	New	1936-2000	Russian research vessel obs.	2.03 M	664 MB
Auxiliary	New	1946-84	Japanese Whaling Ship Data	20.6 K	5.72 MB
Auxiliary	Repl.	1990-97	PMEL/JAMSTEC buoy data	21.4 M	6.19 GB
Auxiliary	New	1990-98	COAPS research vessel obs.	57.7 K	28.0 MB
Auxiliary	New	1878-94	US Marine Met. Journals	1.8 M	888 MB

1. Release 2.4, which will extend ICOADS through 2006, is planned for completion by mid-2007.

2. Addition of ICOADS QC flags is planned, as well as creation of an abbreviated version to resolve technical issues associated with very long supplementary data (C6) lengths.

Table 3 provides examples of the record types currently utilized for data products in terms of the attachment structure. The Storage volume can be minimized in IMMA by omission of attachments that contain only missing data or are not relevant to a given product. For example, the model quality control (C3) information presently is available only for the limited set of VOSclim ships, thus it is unnecessary to include C3 with other products. Similarly, the ship metadata attm (C4) contains metadata selected from WMO-No. 47 (1955-) by the Southampton Oceanography Centre (Kent, et al., 2007a), which currently have been blended only for the periods of 1973-2005.

Table 3. Examples of the IMMA record types currently in use. The notation •/∂ indicates a statically/dynamically included IMMA format component (i.e., the dynamic components are only included if they contain extant data, whereas the static components are always included). For example buoy records included in the ICOADS will never have C4. All these records types may be variable, so the average record sizes are listed (bytes).

<i>Record type utilization</i>	<i>C0 core</i>	<i>C1 icoads</i>	<i>C2 Immt2</i>	<i>C3 model</i>	<i>C4 meta</i>	<i>C5 hist</i>	<i>C6 suppl</i>	<i>Average Rec. size (B)</i>
ICOADS 1998-2005 <sup>1</sup>	•	•	∂		∂		•	327
ICOADS 1973-97 <sup>1</sup>	•	•	∂		∂			213
ICOADS 1784-1972 <sup>1</sup>	•	•	∂					170
VOSclim <sup>2</sup>	•	•	•	•			•	(N/A)
CLIWOC <sup>3</sup>	•						•	528
US Marine Met. Journals <sup>1</sup>	•	•					•	503

1. Data availability: [icoads.noaa.gov](http://icoads.noaa.gov)

2. Data availability: [www.ncdc.noaa.gov/vosclim.html](http://www.ncdc.noaa.gov/vosclim.html)

3. Data availability: [www.ucm.es/info/cliwoc/](http://www.ucm.es/info/cliwoc/) or [icoads.noaa.gov](http://icoads.noaa.gov)

#### 4. Software and documentation

Fortran software (“rdimma0”) to read the IMMA is available within the following website directory: [icoads.noaa.gov/software/](http://icoads.noaa.gov/software/). This software is widely portable, and reads IMMA record types with any legal attachment layout (including core-only records) into a convenient data structure. With minor modifications, the software can also be adjusted to write out IMMA records, once the input data are transformed and loaded into that same data structure, which has proven to be relatively straightforward for many marine data. This self-contained software (which requires only a Fortran compiler) contrasts with BUFR or other Table-Driven Codes (TDCs), which may also require large subroutine libraries for encoding/decoding.

Within the documentation, each IMMA field has a simple alphanumeric identifier (e.g., SST for sea surface temperature, and LAT for latitude). These identifiers, as well as legal ranges for each field and other constraints, are built into the software for ease of use.

The latest version of the attached (Appendix A) IMMA documentation ([imma.pdf](#)) is available within this website directory: [icoads.noaa.gov/e-doc/imma/](http://icoads.noaa.gov/e-doc/imma/). Also stored in the directory are an abridged version of the report ([imma\\_short.pdf](#)), which omits the lengthy background material and provides only the format details contained in supplements C and D of the report, and a technical document ([Pub47\\_IMMA.pdf](#)) providing additional technical details on the blending of the VOS metadata (Kent et al. 2007a).

#### 5. Future plans

For the ETMC (and TT-DMVOS), the maintenance and evolution of the IMMA format in relationship to that of IMMT, plus possible convergence of the two formats, could be an important area of consideration. At present, for example (as noted above) attm C2 (IMMT-2/FM 13) has not been updated (nor a new attm defined) to incorporate changes made effective 1 January 2007 for the IMMT-3. Similarly, no plans been made yet to accommodate changes newly proposed for the IMMT-4 (ETMC-

II/Doc. 3.2). Moreover, the COADS is a large and complex archive, which, in addition to reports from VOS, includes many from buoys and other automated Ocean Data Acquisition System (ODAS). Therefore, any proposals for the IMMA format changes or for possible convergence with the IMMT would need to be thoroughly coordinated and implemented with due consideration of cost and transition issues.

Under the new WMO Information System (WIS), the requirement has been expressed to move all observational GTS traffic (and possibly some other data exchanges) to use TDCs such as BUFR or CREX (ETMC-II/Doc. 3.3, WMO 1995). However, the TDCs are optimized for contemporary and operational data requirements, and the need to store all possible forms of meteorological data leads to a high degree of complexity (and the suitability of TDCs for permanent archival is undemonstrated).

Therefore, considering the success of the IMMA format in the research community and features that appear to make it very attractive for permanent archival, a possible future direction for the JCOMM and the ETMC might be to explore some limited convergence of the IMMA with appropriate features of TDCs (e.g., establish cross-references between the IMMA field names and BUFR/CREX table numbers, and demonstrated record export capability from BUFR/CREX so that modern records can be merged with historical records in ICOADS, thus helping to ensure the continued homogeneity of long-term climate evaluations).

Finalization of the historical atmm (C5) is another important unresolved area. Careful consideration and planning will be needed to develop generalized C5 fields to store a selection of data commonly reported from early ships. This will be beneficial in making original data forms (e.g., Beaufort force numbers, tenths of sky clear or cloudy, and magnetic wind directions) readily available to researchers or to help facilitate readjustments into standardized units (e.g., a uniform retranslation of wind forces to  $\text{m s}^{-1}$  according to a scale other than WMO 1100).

Table 5 outlines some additional areas for future work, based on recommendations from the most recent MARCDAT Workshop, as well as earlier workshops.

Table 5. Recommendations relating specifically to the IMMA format from MARCDAT-II (Kent, et al., 2007b), which are tracked at: [www.marineclimatology.net](http://www.marineclimatology.net).<sup>1</sup>

<u>Recommendations</u>	<u>Individual observations</u>
21	Integrate datasets to the ICOADS using IMMA attachments: The integration of appropriate datasets into the ICOADS should be accomplished using IMMA attachments to provide ancillary information. Priorities for integration should include: information for the interpretation of cloud information including the solar elevation and the relative lunar illuminance from the Extended Edited Cloud Reports Archive (EECRA), and meteorological reports from research vessels.
33	Expand use of the IMMA attachments for the ICOADS: The role of the IMMA attachments for incorporating quality control, bias, metadata and other non-standard information into the ICOADS should be expanded. Tools and documentation should be developed to guide users in the conversion of datasets into IMMA format. The GHRSSST may be considered as a model for this.
35	User feedback: Need both user survey, and methods to capture the feedbacks and bias assessments (e.g., IMMA attachments) that can be provided by users and applications such as re-analyses.
<u>Recommendations</u>	<u>Gridded datasets</u>
71	ICOADS QC: The existing ICOADS QC should be kept intact. Collect alternative QC information and make available with IMMA attachments.

1. Another important technical issue for the ICOADS is that currently available IMMA data prior to 1998 were implemented using the ICOADS end-user "LMRF" format as input, which contains fewer fields than the IMMA, thus, it lacks supplementary (original input) data (ref. Table 3). For the ICOADS, a more inclusive conversion to IMMA, using the full-length production "LMR" format, is needed when resources become available. The LMR format contains important supplementary data, which should be tapped for fields not defined in LMRF but now available in the IMMA (e.g., sea ice fields), or for fields planned for availability in the IMMA when the historical atmm is finalized (e.g., Beaufort wind force numbers).

**References:**

- Diaz, H., C. Folland, T. Manabe, D. Parker, R. Reynolds, and S. Woodruff, 2002: Workshop on advances in the use of historical marine climate data. *WMO Bulletin*, **51**(4), 377–380.
- García-Herrera, R., P. Jones, D. Wheeler, G. Können, and M.R. Prieto (Guest Eds.), 2005: CLIWOC: Climatology of the World's Oceans, 1750-1850 (Special Issue). *Clim. Change*, **73**, 1-194.
- Kent, E.C., S.D. Woodruff and D.I. Berry, 2007a: WMO Publication No. 47 metadata and an assessment of observation heights in ICOADS. *J. Atmos. Oceanic Technol.*, **24**, 214-234.
- Kent, E., S. Woodruff, N. Rayner, T. Arbetter, C. Folland, F. Koek, D. Parker, R. Reynolds, R. Saunders, V. Smolyanitsky, S. Worley, and T. Yoshida, 2007b: Second International Workshop on Advances in the Use of Historical Marine Climate Data. *Bull. Amer. Meteor. Soc.* (in press).
- JCOMM, 2000: Subgroup on Marine Climatology Eighth Session, Asheville, NC, USA, 10 to 14 April 2000, Final Report. JCOMM Meeting Report No. 2.
- JCOMM, 2004: *Expert Team on Marine Climatology, First Session, Gdynia, Poland, 7-10 July 2004, Final Report*. JCOMM Meeting Report No. 32.
- JCOMM, 2007: JCOMM Data Management Programme Area Coordination Group (DMCG), Second Session, Geneva, Switzerland, 10-12 October 2006, Summary Report. JCOMM Meeting Report No. 43.
- Parker, D., E. Kent, S. Woodruff, D. Dehenuw, D.E. Harrison, T. Manabe, M. Meitus, V. Swail, and S. Worley, 2004: Second JCOMM Workshop on Advances in Marine Climatology. *WMO Bulletin*, **53**(2), 157-159.
- WMO, 1955–: *International List of Selected, Supplementary and Auxiliary Ships*. WMO–No. 47, Geneva, Switzerland, (serial publication; recently annual; Editions prior to 1970 were titled *International List of Selected and Supplementary Ships*).
- WMO, 1995: *Manual on Codes*. WMO–No.306, Geneva, Switzerland (Vol. I.1: 1995 Ed. including Suppl. through No. 3 (VIII.2001); Vol. I.2: 2001 Ed.).
- WMO, 2001: *Joint WMO/IOC Technical Commission for Oceanography and Marine Meteorology, First Session, Akureyri, 19-29 June 2001, Abridged Final Report with Resolutions and Recommendations*. WMO–No.931, Geneva, Switzerland.
- Woodruff, S. D., H.F. Diaz, S.J. Worley, R.W. Reynolds, and S.J. Lubker, 2005: Early ship observational data and ICOADS. *Climatic Change*, **73**, 169–194.
- Worley, S.J., S.D. Woodruff, R.W. Reynolds, S.J. Lubker, and N. Lott, 2005: ICOADS release 2.1 data and products. *Int. J. Climatol.*, **25**, 823–842.

**APPENDIX**

**ARCHIVAL OF DATA OTHER THAN IN IMMT FORMAT:  
THE INTERNATIONAL MARITIME METEOROLOGICAL ARCHIVE (IMMA) FORMAT**